The Effect of Child and Adult Avatars on Spatial Perception in Urban-Scale Virtual Environment

1st Jeongmin Lee, 2nd Yuetong Chen
Department of Computing
Goldsmiths, University of london
London, UK

3rd Marco Gillies
Department of Computing
Goldsmiths, University of london
London, UK
m.gillies@gold.ac.uk

4th Xueni Pan
Department of Computing
Goldsmiths, University of london
London, UK
x.pan@gold.ac.uk

***Abstract*—Spatial perception varies significantly between large-scale environments and smaller, object- or room-scale settings. Individuals process and interpret space differently depending on its scale, using their own bodies as reference points. In order to study spatial cognition as it applies to urban environments, it is vital to study it at a suitable scale and in people navigating those environments. Virtual Reality (VR) provides a way to study our spatial cognition in realistic, large-scale urban environments that participants are able to navigate freely. This paper presents a methodology for studying spatial cognition of large-scale virtual urban environments, including three novel tasks that address spatial memory (memory of landmarks and navigation paths) and preferences for urban features (deleting or adding features). This study applies this methodology to investigate the differences in spatial perception, memory, and layout preferences between people embodied in child and adult avatars. This is an interesting problem, as it opens up the possibility of allowing adults to understand how children experience urban environments differently. Using Virtual Reality, participants experienced an urban-scale environment while embodied in either a child or an adult in a full-body avatar and then undertook our spatial memory and preference tasks. The study showed no difference in the memory tasks, though the participants in child avatars deleted more features. This suggests that there is no evidence that embodiment as a child affects spatial cognition but that it might allow participants to detect more problems with an environment.**

***Index Terms*—Spatial cognition, spatial perception, spatial memory, body size, virtual avatar, embodiment.**

I. Introduction

Human spatial perception and cognition, how we understand the spaces we inhabit, is a key element in disciplines such as architecture, engineering, and urban design. This is not simply a matter of passive perception, but of actively inhabiting a space: moving around it and interacting with it in an embodied way. This can be a difficult subject to study in the physical world, due to the costs of building physical spaces (particularly large-scale ones). Therefore, most existing test of spatial memory are limited to verbal description or 2D drawings [1]. However, Virtual Reality (VR) can provide a cost effective mechanism for the study of spatial perception and cognition. Yet, the embodied interactions focusing on spatial abilities are still comparatively under-researched fields [2]. While some studies have addressed object-scale or room-scale spatial cognition, others investigated larger-scale spatial contexts [3], [4]. Furthermore, the methodologies required to study spaces of varying scales differ considerably. This is due to the activation of different brain regions, such as the parietal cortex when interacting with small-scale objects [5] and the hippocampus and temporal cortex in the context of more expansive environments [6], [7]. Environmental space stands out as it goes beyond the human scale, covering larger areas like cities [8]. Our interactions within this space predominantly rely on movement, with our body acting as a central reference point. Similarly, in Virtual Environments (VEs) that utilise full-body embodiment avatars, the avatar's movements will significantly impact our comprehension and interaction with the large digital space.

In order to address the above issues, we conducted an experiment with 18 participants where they were asked to navigate from home to school, while embodied in either a virtual child or adult (see Figure 1). Our main objective is to understand how embodying different avatars, whether as a child or an adult, influences our spatial perception, memory, and layout preferences, potentially due to variations in body size and walking speed between the avatars. The results of this VR experiment are expected to impact future applications across several domains. For researchers who turn to VEs as testing grounds, the ability to control and minimise extraneous variables in VR offers higher precision in study outcomes. On the creative front, for critical realisation, if avatar attributes influence the perception of space in VR, then intended designs or narratives could be misinterpreted. This is especially salient for designers sculpting experiences in virtual spaces, and acknowledging these differences is crucial. Further implications extend to architecture and urban design. This influence isn't limited to physical dimensions; other subtle aspects tied to avatar embodiment might also weigh in. As VR continues its ascendant trajectory across industries, this nuanced understanding of how we perceive virtual spaces—primarily through different avatars—becomes even more essential.

II. Related work

A. Spatial Cognition in Large Space

A classification of psychological spaces proposed by Montello in 1993 suggested four categories based on the relative

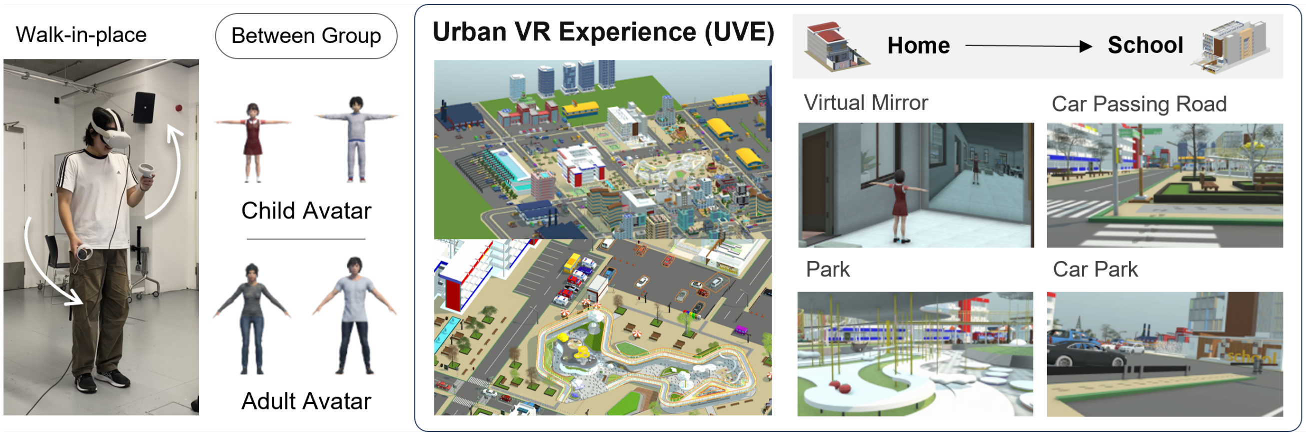

Fig. 1. A total of 18 Participants embodied in a gender-matched adult or child virtual body and navigated from home to school.

projective size of the person's body and the associated actions: figural, vista, environmental, and geographical [8]. Each category hinges on how a person perceives the space in relation to their body and their interaction within it: figural spaces are smaller than a person, allowing for direct manipulation of objects; vista spaces, like rooms, can still be visually grasped from a single viewpoint without needing to move. In contrast, environmental spaces, such as buildings or cities, are so expansive that one needs to move around to gather information. Lastly, geographical spaces are vast or intricate that they are best understood through maps or similar representations.

The perception of space is an example of embodied cognition, as our body influences our mental understanding of space. For instance, a child might perceive objects as more significant due to their relatively smaller size. The visual perspective, determined by the height of our eyes, further affects how distances are perceived. As we move, the somatosensory feedback - the physical effort feedback our body offers - influences our perception of covered distances. Furthermore, the concept of affordances, introduced by J.J. Gibson in 1979, highlights how our environment offers us action opportunities based on our body size [9]. Such complex interplay of body size and spatial perception has been discussed in depth by researchers like Proffitt (2006) [10] and Witt et al. (2004) [11] among others. Their work highlights how our body acts as a reference point, grounding our spatial experiences and shaping our interactions with the surrounding world.

Spatial memory plays a key role in how we perceive and navigate space. Landmarks, often distinct or notable features in a setting, function as mnemonic anchors in our spatial memory [12]. They can provide reference points within a vast space, enabling us to orient ourselves and determine our relative position. When we encounter a known landmark, it triggers spatial memories, aiding in both our current navigation and future recall of that space.

Urban spaces shape our experiences. But how we perceive these spaces can be influenced by our viewing angle. Montello (2014) [13] emphasises the crucial role of "legibility": a legible space enables users to navigate it more efficiently and feel a sense of comfort within it. However, Appleyard (1970) [14] found that different viewing heights could alter one's focus on specific details within a space. Taller individuals might capture wider vistas and overarching urban patterns, while shorter ones might concentrate on immediate surroundings or ground-level intricacies. Such variations suggest that urban design preferences might shift according to these perceptual differences. Consequently, understanding and accounting for these unique perceptual experiences are vital for comprehensive urban planning.

B. Virtual Avatars and Embodiment

VR has the unique capacity to facilitate a first-person perspective via a digital body, and through visual-motor synchrony it generates the illusion that this virtual body is the origin of the users' sensations. Intriguingly, the illusion of embodiment remains resilient even when the avatar differs from the user's age, size, gender, or racial background [15], [16]. In the expansive realm of VEs, the scale of an avatar becomes a defining factor in how users perceive and interact with their digital surroundings. As illustrated by Van der Hoort et al. (2011) [17], there's a direct relationship between the perceived size of one's virtual body and how they gauge distances and object sizes. Participants were made to experience either larger or smaller virtual bodies which then led to notable changes in their perception of space and object size in the VE. The study emphasises the idea that the perceived size of one's virtual body, whether scaled up or down proportionally, can serve as a yardstick for assessing the broader virtual world. Thus, the embodiment in this space is not confined to just personal boundaries but extends to influence our broader spatial perceptions [17]. Kilteni et al. (2012) [18] investigated how participants integrate extremely extended arms into their self-image. Utilising multisensory stimuli in alignment with these modifications, it was found that users could integrate these abnormal proportions into their virtual body schema to a certain extent. Banakou et al. (2013) [19] put adults in the body of a 4-year-old child and also as an adult of the same height. Both scenarios resulted in a pronounced body-ownership illusion, with a more significant size overestimation of objects in the child embodiment. Furthermore, an implicit association test revealed quicker associations of self with child-like attributes in the child's condition. A more recent study by Zhang et al. [20] found that participants would choose to scale a children's chair smaller, and an adult's chair bigger, when perceiving the environment from the eye-level of a 2-year-old as compared to an adult. Collectively, these studies emphasise the dynamic relationship between avatar scale and spatial perceptions. Some avatar changes can significantly influence our perceptions, while others may not. Nonetheless, the research landscape has been largely confined to indoor, room-sized virtual settings, suggesting the possibility of different outcomes in expansive or varied virtual spaces.

C. Research Questions and Hypotheses

So far, most VR embodied studies have been conducted in smaller, room-sized virtual settings. Limited studies have explored the nexus between avatar embodiment and spatial cognition, particularly within the context of urban-scale environments. This gap holds particular significance as urban-scale settings in VEs can offer insights into how individuals navigate, perceive, and interact within large virtual cities or landscapes, providing useful guidelines for urban planning, architectural design, and even transportation modeling in virtual scenarios. The potential benefits of such research could pave the way for more intuitive virtual urban environments, better-designed virtual spaces conducive to user navigation, and a deeper comprehension of how embodiment in virtual avatars influences our perception of expansive spaces. Our core research question therefore is: *Is the perception of large environmental urban spaces in VR influenced by the type of avatar (child versus adult) due to their different characteristics?*

We hypothesise that the type of avatar would influence participants' spatial memory, with the following two hypotheses:

H1: Child avatars, because of their smaller body size, may recall landmarks' positions, facing directions, and sizes less accurately than adult avatars.

H2: When adults embody a child avatar, which is different from their accustomed real-world body size, they might experience a decline in their route recall accuracy, and be less able to remember their navigation routes.

We also hypothesise that in a VR setting, the type of avatar body would influence participants' preferences regarding urban layouts, with our hypothesis formed as:

H3: Preferences regarding the design and functionality of virtual urban spaces might vary based on the avatar's eye level. The shift in perspective position could influence their perception and, consequently, their preferences for the layout design.

III. Experimental Design

The study was designed as a between-subjects experiment to minimise bias. Specifically, each participant was to be exposed to only one condition – a child or adult avatar – to preclude any influence on subsequent tasks related to wayfinding and memory.

The experimental design includes three questionnaires and two VR experiences, executed in the following sequence: pre-questionnaire, an urban-scale VR experience (UVE), mid-questionnaire, a set of VR spatial tasks (VSTs), and post-questionnaire. The questionnaires were conducted via Microsoft Forms. The VR applications for the experiment were developed using Unity Engine 2019.3.27f1 version, and the VSTs consisted of three independent scenarios developed as measurements.

A. Urban VR Experience (UVE)

In the urban VR environment, the primary objective of participants is to navigate around the urban-scale VE from home to school. The key spatial elements along this route include the home, a small playground, a traffic light intersection, a public park, a parking facility, and the school. Participants are tasked with reaching the school while maintaining situational awareness. Although not mandatory, directional markings (footprints) on the ground are provided to assist participants in route finding.

When the participants first enter the scene, they are positioned in front of the mirror, allowing them to view their avatar's full-body reflection. This is intended to increase the level of immersion and enhance the subjective sense of body ownership within the virtual setting [21]. It is designed for participants to take a moment to observe themselves with their virtual reflections in the virtual mirror and practice manipulating the virtual fingers, arms, and navigating using the controllers. Additionally, the scene includes potentially hazardous situations, such as moving vehicles and elevated platforms, to induce physiological arousal in participants when they encounter them. Once participants successfully arrive at the school, their experience in UVE will be considered complete.

Throughout this experience, participants navigate through a first-person perspective with the inclusion of corresponding head rotations. Controllers are used for the two types of locomotion. Participants should swing their arms in a walk-in-place motion to simulate normal walking while holding both left and right controllers. For more incremental movements, the joysticks on both the left and right controllers are assigned for continuous movement and continuous turn, respectively.

Moreover, arm-swinging locomotion - facilitated by a walk-in-place mechanism - plays a critical role in supplementing the augmentation of participants' spatial updating and cognitive mapping capabilities. While this framework does not fully emulate real-world walking conditions – due to the lack of information on the lower body parts engagement in the absence of a treadmill and limited physical space: unable to walk physically urban-scale size – participants still acquire partial body-based information. Moreover, they gain access to an ambient motion array and optical flow [9], enriching the environmental information obtained through locomotion using their avatar's body as a reference point.

Characteristics of the avatar, such as age, appearance, eye level, and walking speed, are assumed to influence participants' spatial experiences, particularly in terms of perceived scale and memory retention. Specifically, adjustments are made to the eye level positions, and movement speeds to distinguish perceptual experiences between child and adult avatars. Regarding eye level, the y-axis value for the child avatar is set 30 units lower than an adult. Regarding locomotion speeds, different settings are employed for continuous movement and walk-in-place: for continuous movement, the child avatar is set to a speed of 0.5, while the adult avatar is set to 0.8; for walk-in-place, the speed is at 1.8 for the child and 3 for the adult.

A pre-configured "Virtual Reality Inverse Kinematics (VRIK)" is employed to generate a more realistic full-body movement of the avatar. The fundamental principle of inverse calculation is based on the positions of three pivotal points – the head, left hand, and right hand – from which the system is able to predict and simulate the leg position and movement without requiring input from the lower parts of the body.

Hand animation functionality is integrated using the "Auto Hand" package, with manual configuration of hand poses. Individual control is enabled for the thumb, index finger, and remaining three fingers, allowing the bending motion to facilitate interaction with the environmental elements within the scene. This motion can be initiated through button presses on the controller.

B. VR Spatial Tasks (VSTs)

Spatial thinking is crucial for visualization, facilitating spatial reasoning and problem-solving. Traditional approaches to studying spatial reasoning often employ paper-pencil or computerized tasks. However, these 2D methods can't fully capture the richness of our 3D interactions in real-world spaces [22]. Paper-pencil tasks, although straightforward, don't reflect the full depth of spatial experiences. VEs offer a solution. They overcome the constraints of 2D evaluations by simulating a more immersive 3D space. For instance, a study assessing mental rotation in VR found that participants performed better with 3D stimuli in VR than with 2D images [23]. Using this as a benchmark, a task encompassing an environmental scale with 3D stimuli for assessing the spatial memories and the preferences was designed. This points to the value of using VEs for a more accurate spatial memory assessment.

Participants are required to complete three spatial tasks related to their prior experience in the UVE. Spatial Task 1 (ST-1) focuses on spatial memory related to landmarks, Spatial Task 2 (ST-2) evaluates spatial memory in terms of navigation, and Spatial Task 3 (ST-3) involves redesigning the urban layout. A design illustration is shown in Figure 2. Unlike the UVE, the VSTs only feature hand or controller representations instead of a full-body avatar. While the tasks are still experienced from a first-person perspective, the replica of the urban environment is presented as a miniature, with a bird's-eye view. Participants have the flexibility to adjust the scale of this model, with scaling options ranging from 0.5 to 3 - a miniature size to an almost immersive, near-life-sized view, respectively.

1) Spatial Task 1(ST-1): Spatial Memory (Landmarks): In ST-1, participants encounter five objects on the table: landmark 1, landmark 2, home, school, and park. The task aims to assess their ability to accurately remember and position these landmarks based on their previous experience navigating the virtual urban environment. Participants can use their controllers to grasp and place these landmarks, with the option to adjust their orientation through rotation and modify their size by scaling them up or down. At the same time, landmarks can be positioned outside the boundaries of the urban model, where they remain static. However, gravity will be enacted when they are situated slightly above the model, and they will descend to the ground. Once participants are content with the placement, scale, and orientation of all landmarks, the task can be concluded by pressing a button located behind them.

Landmarks can be moved or rotated using both controllers with a button press. Specifically, the trigger or grip button simulates the participants' fingers, providing an intuitive feeling as though they are directly using their own actual hands. By holding a component using both hands and pressing the trigger or grip button, participants can adjust its scale through zooming actions, either enlarging or reducing its size.

2) Spatial Task 2(ST-2): Spatial Memory (Navigation): In ST-2, the objective for participants is to accurately retrace the route they previously navigated in the urban environment. To accomplish this, they use controllers to manipulate a figure that initially appears at a starting point, identified as "home". This task aims to evaluate their spatial memory in terms of navigation. Participants are advised to avoid mistakes or redundant movements. They can release the figure if they need to pause to recollect their memory of the route. When they are confident, the task can be concluded by pressing a button located behind them. Participants hover their controller over the figure, triggering a colour change to dark blue to begin moving it. This colour change indicates that the figure is ready to be gripped. Participants must press and hold the grip button on their controller to grab and move the figure. Specifically, the figure is designed to move only within the horizontal plane of the model, restricted to the x and z axes. This eliminates the possibility of height adjustments (as flying was impossible in the UVE) and is implemented to reduce the risk of participants accidentally dropping the figure. The position values are continuously recorded at an interval of 0.5 seconds, considering the scale of the environment, to capture the route the participant has drawn with the figure.

3) Spatial Task 3 (ST-3): Urban Layout Preference: In ST-3, the aim is to probe participants' urban design preferences influenced by their different experiences with different avatars (child and adult) in the UVE. This scenario is designed to identify how different virtual experiences might influence one's approach to urban design and functionality.

Participants can remove up to five of 19 district blocks from the original model. Additionally, they are provided with 24 elements, with each component available in a quantity of three.

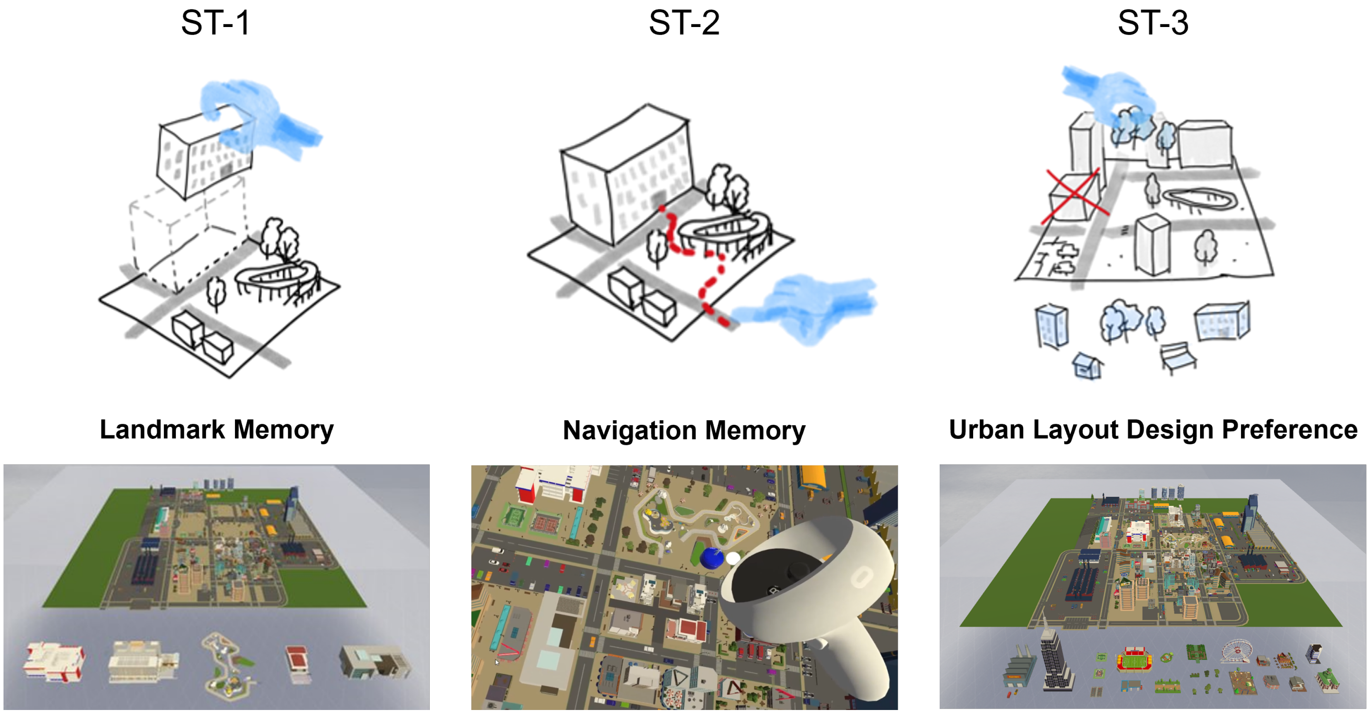

Fig. 2. The three VR spatial tasks: design (up) and final implementation in VR (down). ST-1 required participants to position the absent landmarks in the correct positions with accurate scales and orientations. ST-2 required them to draw the path they had taken in the first VR navigation task. In ST-3, participants were asked to remove and add elements to the urban space based on their own preferences.

The act of adding elements is not constrained. These elements have been classified into four categories: infrastructure, nature, architecture, and entertainment.

Aligned with ST-1's logic, the gravity functions remain available. However, while ST-1 assessed participants' spatial skills through scaling transformations, components in ST-3 are arranged explicitly to human scale; thus, modifying any size adjustments is infeasible. The names of the elements are documented via a JSON file, including both the elements participants choose to remove and the ones they add.

C. Participants

The eligibility criteria for the study included being 18 or above and willing to explore a virtual urban environment. A total of 18 participants (9 male, 9 female) were recruited. However, due to the issues encountered during the data exporting process, two participants (one male and one female) were excluded. Two participants out of the 18 had no prior VR experience. The study was approved by our ethics committee.

D. Hardware

Both UVE and VSTs were displayed using an Oculus Quest 2 head-mounted display (HMD) connected to the desktop with a 3-meter link cable. The graphics were rendered on a Windows 11 desktop powered by an AMD Ryzen 9 5900X 12-Core processor with the NVIDIA GeForce RTX 3070 GPU graphic processing. For interaction, participants used the default Oculus touch controllers in each hand.

E. Procedure

The experiment was conducted at the our VR research lab, located on the campus of Goldsmiths University in London. On arrival, participants received an information sheet and were asked to sign a consent form. Once these initial steps were accomplished, each participant completed a five-phase procedure as follows: (1) pre-questionnaire; (2) urban VR experience; (3) mid-questionnaire; (4) VR spatial tasks; and (5) post-questionnaire.

Participants were randomly allocated either a child or adult avatar, which is gender matched: a boy or man for male participants and a girl or a woman for female participants.

The initial phase involved the pre-questionnaire. For consistency, participants were handed an instruction sheet detailing the objectives within the UVE and the controller's functionalities. Participants were encouraged to ask any questions for clarification. Before launching the VR session, they were guided to adjust the lens distance for optimal vision and then stand in a position. Following this, the first VR experience (UVE) began. Upon entering the scene, participants were asked about the avatar's eye level accuracy by referencing a virtual mirror. This scenario was set with a 10-minute limit, yet all participants completed it within this time frame. Following the first scenario, they filled out the mid-questionnaire.

After this, they were given an instruction sheet detailing the spatial tasks. Beginning with the general overview, then received specific instructions for ST-1. Once ST-1 was completed, participants were asked to remove the HMD and review the corresponding guidelines for subsequent tasks. This process was repeated until the final task (ST-3) was conlcuded. A comprehensive description of each VR scenario is provided in the 'Experiment Design' section. After completing the second VR experience, participants were asked to fill out the post-questionnaire, marking the end of their participation. On average, each session lasted about 45 minutes.

IV. Results

Although questionnaire data was collected during our experiment, in this paper we focus on the result from the spatial tasks developed specifically for this study. Outcomes from the VSTs were analysed with an independent sample t-test using IBM SPSS version 27.

A. ST-1: Landmarks

ST-1 was designed to assess participants' spatial memory of landmarks. The task required participants to match the initial transform values of five specific landmarks. Three attributes were recorded for each landmark: (1) Position, (2) rotation, and (3) scale, with x, y, and z values documented for each in a JSON format. Each attribute had its distinct scoring criteria, ranging from a minimum of 0 to a maximum of 5 points (Figure 3). Initially, individual scores were calculated for every attribute of each landmark. Subsequently, the mean of these scores, derived from the five landmarks' attributes, was computed distinctly to evaluate participants' skills in placement (1), orientation (2), scale (3), and overall ability (total). The data was then processed, calculated, and visualised using Python within the Spyder IDE.

Fig 3 illustrates the responses from participants for the ST-1 and the explanation of the visualisation. These are orthographic projections, seen from a top view of the result. The direction in which the y-axis faces (denoted by the triangle) signifies the front direction of the landmark. A (r) symbol indicates the landmark is oriented in the correct direction, whereas a (s) denotes a rotation of ±180 degrees from the proper alignment. Accompanying scale scores, the symbols ((0), (-), (+)) specify whether the provided scale in the response is precisely aligned, smaller, or larger than the original landmark's scale, respectively.

Test for normally in SPSS shows that our residuals error from a linear regression are normally distributed (VIF <10 for all three measurements). An independent t-test found no significant differences were found for position or scale (Position: $t(14) = .446$, $p = .662$; Scale: $t(14) = .712$, $p = .276$). However, there is a marginal significant difference for rotation score ($t(14) = 2.003$, $p = .065$), with participants in the child group scoring higher than the adult group (Child Mean = 4.33, Adult Mean = 3.98).

B. ST-2: Navigation

The evaluation for ST-2 was focused on comparing the route taken from UVE with the drawn path from ST-2. Two primary metrics were employed: the Mean Squared Error (MSE) and Ending Point Error (EPE). MSE quantifies the average squared differences between the corresponding coordinates of two routes, offering a precise measure of their similarity. In contrast, EPE assesses the disparity between the endpoint in the UVE and the reproduced endpoint in ST-2.

Since there was a difference in the time durations between the two data sets, ensuring synchronous alignment was crucial. To achieve this, linear interpolation was utilised, matching the timestamps of the ST-2 dataset with those of the UVE.

Spatial scale consistency was another critical consideration. Distance between specific reference points (home and school) within each dataset was used to ensure this. A scaling ratio was then derived from these measurements to ensure both datasets were spatially normalised.

After normalisation of both timesteps and scale, the primary comparison was executed using MSE. A lower MSE indicates closer alignment between the routes, while a higher MSE suggests greater discrepancies between them. This computation was implemented via Python's *scipy.interpolate* library. Although initial attempts were made to evaluate the turning tendency of the routes, it was found that the MSE provided a more accurate representation, leading to its selection as the preferred metric. Furthermore, EPE was computed using the Euclidean distance between the two routes endpoints. For visualisation, both data sets were adjusted so their starting points coincided. These paths were then rendered on a 2D grid, marked distinctly at the beginning with a circle and at the end with a square. The EPE was represented as a pink dotted circle centred on the ST-2's endpoint. The extent of this circle, up to the UVE's endpoint, clearly illustrates the distance between the two concluding points.

Both MSE and EPE passed the normality for residual error after linear regression test (VIF$<$10). An independent t-test found no differences between the two conditions for both measurements (MSE: $t(14) = .816$, $p = .428$; EPE: $t(14) = -.323$, $p = .752$), indicating no significant difference in their navigation pattern between child and adult conditions.

C. ST-3: Urban Layout Preference

The ST-3 evaluation had no specific numeric answers against which participants could be assessed. Instead, obtaining the names of both removed blocks and added components made it possible to discern participants' tendencies regarding their preferences.For the analysis, individual removed blocks, categorised added components, and determined totals for both.

We then looked at the number of objects replaced. Our data passed the normality check for residual error after linear regression confirms (VIF$<$10). An independent t-test found a significant difference between the two conditions ($t(14) = 2.175$, $p = .047$), suggesting that participants in the child condition replaced more items than the adult one (Child: Mean= 4.13; Adult Mean = 2.75).

V. Discussion

Our first hypothesis (**H1**) examines the participants' ability in landmark recall. We hypothesised that those who embodied in an adult body would be able to perform better in recalling

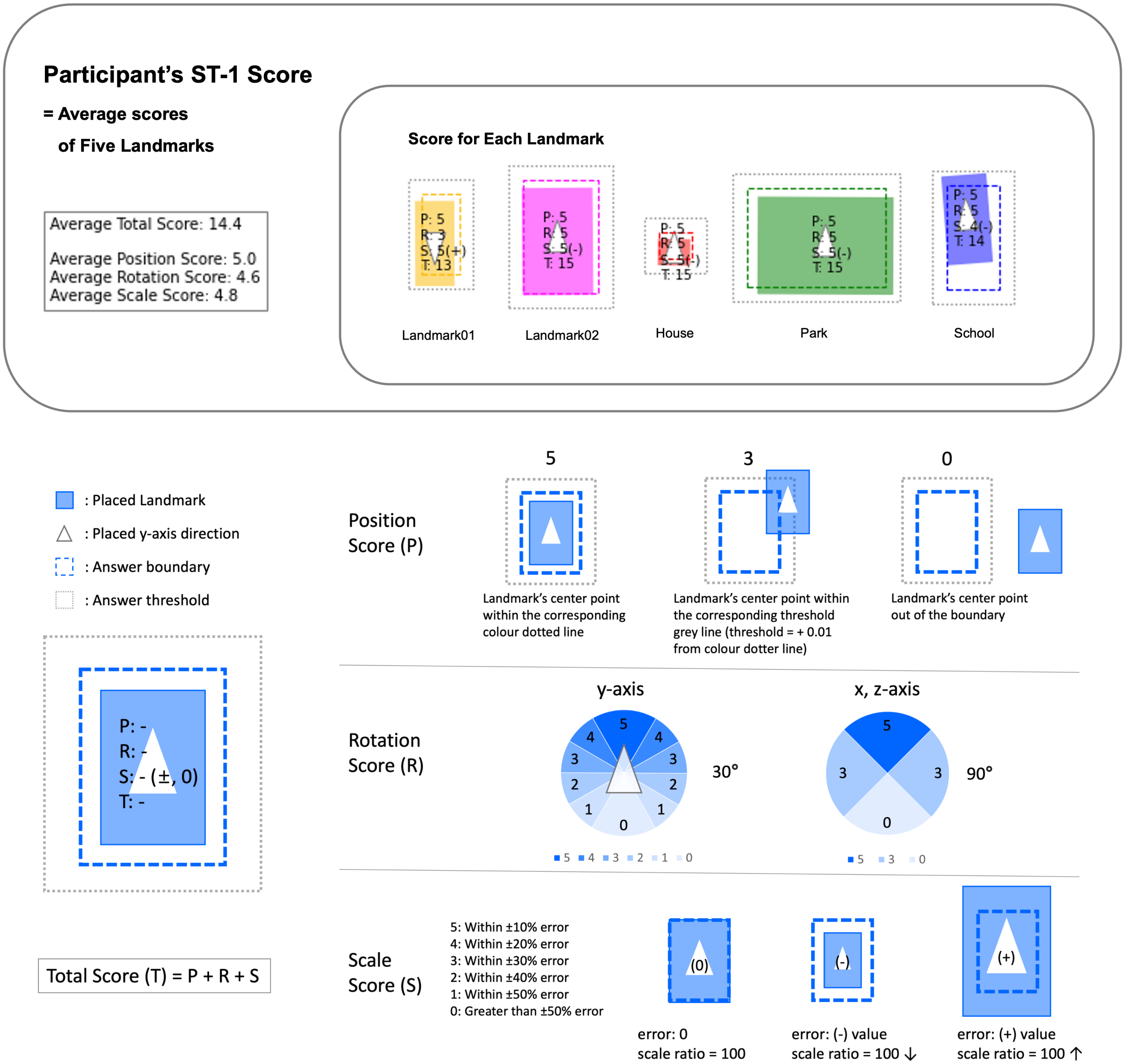

Fig. 3. ST-1 Top: average score and an example score from one participant; Bottom: Scoring criteria for position, rotation, and scale

landmark's position, direction, and scale, due to their virtual body representation being more similar to their real-world ones. However, no differences was found in participants' ability to recall landmarks position and scale. This is inconsistent with the study by Banakou et al., which showed that people who embodied in a child avatar showed an overestimation of the object size [19]. This might be because, as outlined before, based on Montello's definition of scale [8], the object is included as a figural space, while our urban-scale VR is an environmental space, and their mechanism in terms of spatial reasoning differs. Such findings suggest that factors other than avatar size might be instrumental in shaping perceptions of landmark size within virtual realms. Interestingly, there was a trend indicating that participants who had a virtual body of a child preformed better in their rotation abilities. However, this could simply because participants who underwent the child condition were better at rotation abilities. This is a limitation in our study which has a very small sample size. Future work could measure spatial skills in the pre-questionnaires to ensure both conditions have a similar distribution of this ability among participants.

Our **H2** hypothesise that adults taking on a child avatar might face challenges in route recall due to the discrepancy between the avatar's size and their actual body size. In evaluating this, the study employed two metrics: Mean Squared Error (MSE) for gauging overall route recall precision and Ending Point Error (EPE) to measure the accuracy of recalling the route's conclusion. An independent sample t-test confirmed that differences in these measures were not statistically significant. This suggests that factors beyond avatar size play a role in influencing navigational recall within VEs.

H3 proposed that avatar eye level could influence design preferences in virtual urban spaces. In the ST-2 evaluation, rather than numeric scores, preferences were inferred by observing which blocks participants added or removed. From the data, a key difference emerged in the 'removed total' category. Specifically, participants using child avatars removed an average of 4.13 blocks, while those with adult avatars removed an average of 2.75 blocks. This difference was statistically significant, with a t-test p-value of 0.047. Even though the average number of total added components was 5.5 for child avatars and 8 for adult avatars, this difference was not statistically significant. However, this trend was only observed in the 'removed total' category. Other evaluated metrics, such as the number of individual blocks added or removed, exhibited consistency across avatar types, suggesting that while avatar height might influence some aspects of design preferences, its overall impact might be limited. It opens up the intriguing possibility that participants embodied as children might be noticing problems with an environment which adults would not notice. This needs to be investigated in future studies.

VI. Conclusions

This paper has presented a methodology for studying spatial perception and cognition in urban-scale environments using VR. The focus was on three spatial tasks, which assessed different aspects of spatial memory and participants' preferences for spatial layout. This method was applied to a study of differences in spatial perception between participants embodied in child or adult avatars while navigating a virtual urban environment. The study showed that child and adult avatars differed very little in spatial memory. They did, however, differ significantly in their preferences for layout. This is a very interesting result, as it opens up the possibility of using child avatars to help urban planning professionals take a more child-centered view of urban design. The use of child avatars would not impact their understanding of a space, but it might highlight problems that might not be visible at an adult scale (more work is needed to confirm this). However, perhaps the most important contribution of this paper is the overall method that allows us to study spatial cognition of urban space using a flexible VE and a number of cognitive tasks. We hope that this method can be applied to a wide variety of cognitive studies in the future.

References

[1] U. Diaz-Orueta, B. M. Rogers, A. Blanco-Campal, and T. Burke, "The challenge of neuropsychological assessment of visual/visuo-spatial memory: a critical, historical review, and lessons for the present and future," *Frontiers in Psychology*, vol. 13, p. 962025, 2022.

[2] S. Lee-Cultura and M. Giannakos, "Embodied Interaction and Spatial Skills: A Systematic Review of Empirical Studies," 2020.

[3] G. Franz, M. Von Der Heyde, and H. H. Bülthoff, "An empirical approach to the experience of architectural space in virtual reality-exploring relations between features and affective appraisals of rectangular indoor spaces," in *Automation in Construction*, vol. 14, no. 2 SPEC. ISS., 2005.

[4] S. H. Cha, C. Koo, T. W. Kim, and T. Hong, "Spatial perception of ceiling height and type variation in immersive virtual environments," *Building and Environment*, vol. 163, 2019.

[5] J. Zacks, B. Rypma, J. D. Gabrieli, B. Tversky, and G. H. Glover, "Imagined transformations of bodies: An fMRI investigation," *Neuropsychologia*, vol. 37, no. 9, 1999.

[6] E. I. Moser, E. Kropff, and M. B. Moser, "Place cells, grid cells, and the brain's spatial representation system," 2008.

[7] A. D. Redish and A. Ekstrom, "Hippocampus and related areas: What the place cell literature tells us about cognitive maps in rats and humans." in *Handbook of spatial cognition.*, 2012.

[8] D. R. Montello, "Scale and multiple psychologies of space," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 716 LNCS, 1993.

[9] J. J. Gibson, *The ecological approach to visual perception*. Lawrence Erlbaum Associates, 1979.

[10] D. R. Proffitt, "Embodied Perception and the Economy of Action," *Perspectives on Psychological Science*, vol. 1, no. 2, 2006.

[11] J. K. Witt, D. R. Proffitt, and W. Epstein, "Perceiving distance: A role of effort and intent," *Perception*, vol. 33, no. 5, 2004.

[12] K. Lynch, *The image of the city*. MIT press, 1960.

[13] D. R. Montello, "Spatial cognition and architectural space: Research perspectives," 2014.

[14] D. Appleyard, "Styles and methods of structuring a city," *Environment and Behavior*, vol. 2, no. 1, 1970.

[15] D. Banakou, P. D. Hanumanthu, and M. Slater, "Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias," *Frontiers in Human Neuroscience*, vol. 10, no. NOV2016, 2016.

[16] M. Gonzalez-Franco and T. C. Peck, "Avatar embodiment. Towards a standardized questionnaire," 2018.

[17] B. van der Hoort, A. Guterstam, and H. H. Ehrsson, "Being barbie: The size of one's own body determines the perceived size of the world," *PLoS ONE*, vol. 6, no. 5, 2011.

[18] K. Kilteni, J. M. Normand, M. V. Sanchez-Vives, and M. Slater, "Extending body space in immersive virtual reality: A very long arm illusion," *PLoS ONE*, vol. 7, no. 7, 2012.

[19] D. Banakou, R. Groten, and M. Slater, "Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 31, 2013.

[20] J. Zhang, Z. Dong, X. Bai, R. W. Lindeman, W. He, and T. Piumsomboon, "Augmented perception through spatial scale manipulation in virtual reality for enhanced empathy in design-related tasks," *Frontiers in Virtual Reality*, vol. 3, p. 672537, 2022.

[21] M. González-Franco, D. Pérez-Marcos, B. Spanlang, and M. Slater, "The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment," in *Proceedings - IEEE Virtual Reality*, 2010.

[22] C. He, E. R. Chrastil, and M. Hegarty, "A new psychometric task measuring spatial perspective taking in ambulatory virtual reality," *Frontiers in Virtual Reality*, vol. 3, 2022.

[23] I. Lochhead, N. Hedley, A. Çöltekin, and B. Fisher, "The Immersive Mental Rotations Test: Evaluating Spatial Ability in Virtual Reality," *Frontiers in Virtual Reality*, vol. 3, 2022.